Pergamon

0042-6989(94)00284-3

Vision Res. Vol. 35, No. 14, pp. 1991–2005, 1995

0042-6989/95 $9.50 + 0.00

Detection of Orientationally Multimodal Textures

D. R. T. KEEBLE,*† F. A. A. KINGDOM,†§ B. MOULDEN,‡ M. J. MORGAN*§

Received 19 November 1993; in revised form 25 October 1994

Oriented textures were produced with the use of probability density functions modulated sinusoidally over orientation. Orientational contrast sensitivity functions (OCSFs) for a task involving the discrimination of these patterns from orientationally-random textures were found for several human observers. An inverse Fourier transform of this OCSF yielded a weighting function, or filter, defined over orientation. The weighting function is broad, with a half-height full-width of 34 deg. This orientational filter was able to predict human performance in further discrimination tasks employing a variety of probability density functions over orientation.

Texture Psychophysics Orientation Linear systems

INTRODUCTION

The orientation content of a texture is a highly significant factor in determining whether it will be discriminated or segregated from another texture (Beck, 1966; Beck, Prazdny & Rosenfeld, 1983). In particular, the overall line segment orientation distribution is generally a better predictor of segregation strength than the figural similarity of the micropatterns used to construct the texture. Indeed, human perception of many of the micropattern pairs upon which Julesz (1962) based his conjecture that texture segregation depends on second-order statistics can also be explained in terms of the orientation properties of the textures. Most contemporary models of human texture perception regard the processing of texture properties as an essentially low-level activity [see Bergen (1991) for a useful review]. Such models generally employ banks of oriented filters analogous to the orientationally-selective units known to exist in the primate primary visual cortex (e.g. Landy & Bergen, 1991; Malik & Perona, 1990; Caelli, 1982; Graham, Beck & Sutter, 1992; Bergen & Adelson, 1988). Therefore, in order to probe psychophysically the nature of the integrative mechanism that produces descriptions of texture properties for use by higher levels of the visual system it may well be a good strategy to assess performance in tasks involving textures where the orientational content is easily quantifiable.

There is a long tradition in vision research of employing linear systems techniques in order to understand the nature of the transformation the visual system performs upon an image presented to the retina. The work of Campbell and Robson (1968) on the perception of sinewave gratings at threshold is perhaps the best known example in vision of the use of linear systems methods and, specifically, of Fourier analysis. The application of Fourier methods to the analysis of human performance in visual tasks has concentrated on the representation and interpretation of visual stimuli in the spatial and temporal position/frequency domains, although there have been occasional forays into other domain pairs, such as wavelength and chromatic frequency (e.g. Barlow, 1982). In this paper we seek to apply these ideas to novel territory: the orientation domain.

In general, those workers who have conducted careful psychophysical investigations into the perception of orientationally defined texture patterns, (e.g. Nothdurft, 1985a, b; Landy & Bergen, 1991; De Weerd, Vandenbussche & Orban, 1992), have employed unimodal orientation distributions. That is, the textures that they have used have only one mean direction. Some examples of unimodal textures can be seen in Figs 2(a), 3(a) and 5. Although both Beck (1966) and Julesz (1975) among others have produced patterns constructed from textons in such a fashion that multimodal textures are produced, there has not, to our knowledge, been any attempt to link their multimodal nature with psychophysical results. In spite of the fact that unimodal textures may well be more naturalistic than those with more complicated orientation distributions (see Discussion), the mechanisms that subserve their processing by

*Laboratory for Neuroscience, Department of Pharmacology, University of Edinburgh, Edinburgh EH8 9JZ, Scotland.
†To whom all correspondence should be addressed at: McGill Vision Research Centre, Department of Ophthalmology, 687 Pine Avenue West, H4-14 Montreal, Quebec, Canada H3A 1A1 [*Email* dkeeble@jiffy.vision.mcgill.ca].
‡Department of Psychology, University of Western Australia, Nedlands, Perth, W.A. 6009, Australia.
§Institute of Ophthalmology, University of London, Bath Street, London EC1V 9EL, England.

the human visual system are presumably the same. Knowledge of how texture perception varies with the number of orientation modes should thus provide information about these mechanisms. We have therefore studied texture discrimination tasks using stimuli with a variety of orientation probability density function types and with different numbers of orientation modes.

The key idea is to investigate the properties of the mechanism responsible for the perception of oriented textures by measuring thresholds for tasks involving textures constructed from line elements whose orientations are determined stochastically by a probability density function (pdf) which varies *sinusoidally* with orientation. With such stimuli we can measure an orientational contrast sensitivity function (OCSF). Here the term "contrast" refers to the modulation in the pdf as a function of orientation—not to the conventional luminance contrast. The OCSF is a function of *orientation frequency*, which is proportional to the number of orientation modes. If one assumes that the visual system is responding to the sinusoidal basis functions in a linear fashion, then the OCSF may be regarded as a orientational modulation transfer function (OMTF), in which case its inverse Fourier transform will be a generalized orientation weighting function, and can be used to predict the response of the system to any orientation distribution. The use of textures comprising lines with a variety of orientations helps to ensure that the discrimination task is one which can only be performed effectively by some kind of integration of orientation information over space.

Riley (1981), and Voorhees (1987) have each produced demonstrations involving regions of simple comb-filtered bimodal textures embedded in orientationally random textures. An example of a bimodal comb-filtered texture can be seen in Fig. 4(a). Their images appear to show that segregation in this case is difficult, but they did not conduct rigorous psychophysical investigations to quantify this conclusion. Chubb, Econopouly and Landy (1994) have modelled human psychophysical performance in texture tasks involving stimuli comprising pixels with greylevels independently drawn from histograms with a variety of different shapes. The spirit of their approach is somewhat similar to ours, although the class of stimuli used is completely different.

METHODS

Display

The experiments were performed using an Apple Macintosh IIcx computer attached to a Macintosh 13″ colour display with a frame rate of 66.7 Hz. The greyscale used by the stimulus-generating program was linearized with respect to the human visual system by means of a photodiode with a $V(\lambda)$ filter. The stimuli consisted of bright white lines on a dark background. Normal room lighting provided a veiling luminance of 25 cd/m^2. The lines were generated using the anti-aliasing algorithm of Gupta and Sproull (1981). The display of the images was synchronized to the vertical blanking interval, and timing was controlled by the VIA timer. The pixel resolution at the viewing distance used (57 cm) was 27.2 pixels/deg.

Stimuli

In each experiment, the task assigned to the observers was to discriminate between two arrays of oriented line elements placed side by side. One array comprised lines with random orientations, and the other comprised lines with non-random orientations. The observers were instructed to select the non-random array. Some examples of the individual arrays used are shown in Figs 2, 3, 4 and 5.

The orientations of the lines were drawn from a given probability density function (pdf). The pdfs used are shown in Fig. 1. It is important to realize that the orientation of a straight line element is only defined over π rad, and that 0 rad is equivalent to π rad. The flat line in Fig. 1(a–d) corresponds to the orientationally random arrays. The three sinusoidal pdfs are examples of the distributions used for the initial experiment, Experiment 1, from which the OCSF comes. We define the orientation frequency N, of these sinusoidal pdfs as the number of cycles in π rad, c/π. In this figure we show $N = 1$ and $N = 2$. This nomenclature is useful, because for integral frequencies N is also equal to the number of

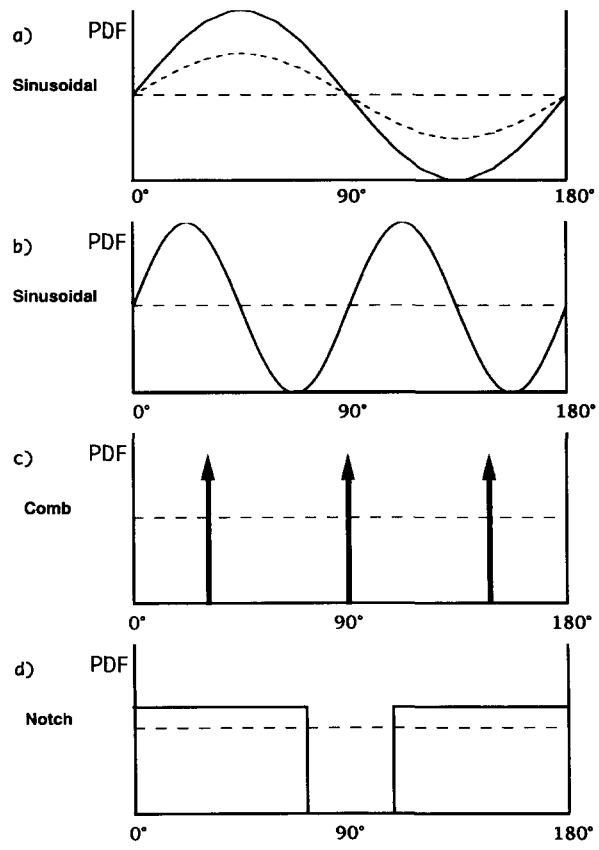

FIGURE 1. The probability density functions (pdfs) employed in the experiments: (a) sinusoidal pdfs at full and half amplitude. Frequency is 1 c/180 deg; (b) sinusoidal pdf with frequency of 2 c/180 deg; (c) comb pdf with $N = 3$; (d) notch pdf with the width of the notch being 35 deg.

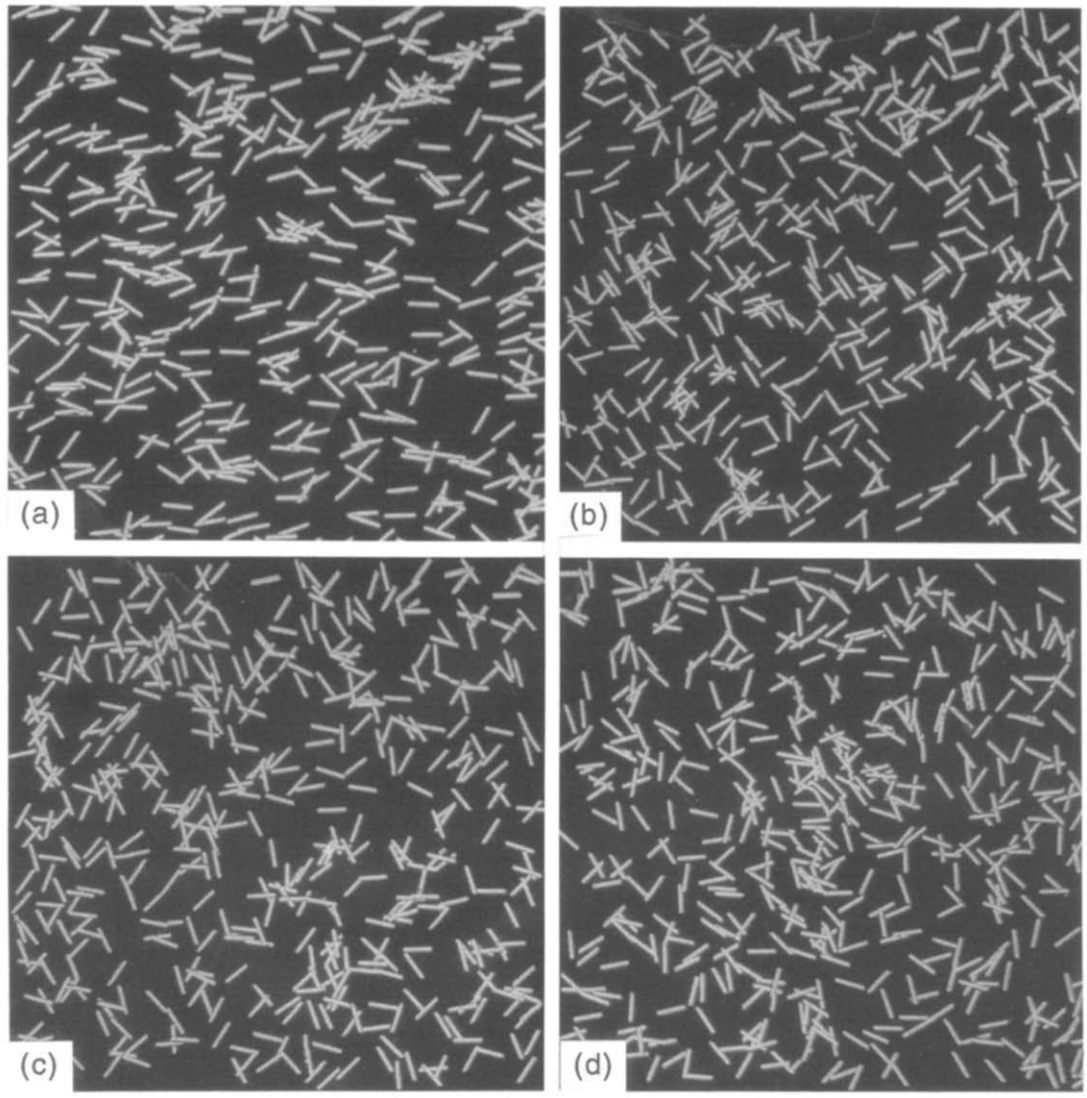

FIGURE 2. Examples of the spatially random stimuli used in Experiment 1. The arrays were produced using sinusoidal pdfs with frequencies of: (a) 1 c/180 deg; (b) 2 c/180 deg and (c) 3 c/180 deg, except for (d), which is orientationally random.

orientation modes. Two out of three of the sinusoids shown here are in fact at maximum amplitude—it is of course not possible to have negative probability densities. As the amplitude decreases and the pdfs become closer to the random case, it is clear that human performance on the task must decline. Hence we take our threshold as being the amplitude of the pdf at a criterion level of discrimination. The phase of the sinusoids with respect to the orientation of the display screen was randomized, in order to prevent subjects from performing the task by attending to the orientation strength in a particular direction. A variety of values of N were used, as can be seen in Fig. 6. We chose a maximum of $N = 3$ to ensure that the form of the orientation pdf was well-sampled.

In the case of N not being integral, the portion of the sinusoid actually used in producing the pdf is randomized. Where $N < 0.5$, it is clear that peak and trough cannot both be present. In the interests of consistency, we define the amplitude as being the amplitude of the original full sinusoid used to define the function used in pdf generation. Because a constant number of lines is used, the pdf is *ipso facto* normalized to a constant. It should be noted that the Monte-Carlo generation of values from the function together with the normalization condition produces a change in the pdf measured in units of probability density. An alternative measure of amplitude which could have been employed would be to take the difference between the actual maxima and minima within each pdf used. This might produce some change in the rate of increase in threshold for $N < 1.0$ seen in Fig. 6, but not in the form of the function.

In the majority of the experiments mentioned here, the line elements were positioned *randomly* within each array. This naturally leads to lines overlaying each other, causing intersections to occur. There is a tradition in the texture literature of regarding such features as particularly salient (Julesz, 1981). We wanted to investigate the possibility that the perception of our textures was affected by the differential occurrence of these intersections, over and above the orientation properties of the image. Accordingly, for the sinusoidal probability density functions we performed an additional experiment employing lines placed upon a grid, in order to

prevent the occurrence of intersections. These displays consisted of 196 lines placed on a 14×14 array. Random positional jitter of $\pm 13.2'$ of visual angle was added in both dimensions so as to minimize the occurrence of collinear structures, which might provide an additional spurious cue. This kind of irregular grid-like texture has been employed extensively by Nothdurft (1985a, b). Examples of the grid-type stimuli used can be seen in Fig. 3. In the case of the random stimuli, 500 lines were used in each array, except for one condition performed by one subject (DK), in which 196 lines were used in order to reproduce the line density of the grid condition.

We wish to test the hypothesis that the orientational function obtained by performing an inverse Fourier transform upon the OCSF data which will be obtained in Experiment 1 represents a linear texture weighting function over orientation. If this is so, then it should be possible to predict behavioural data for discrimination of other orientationally non-uniform textures. To this end we performed Experiments 2 and 3. In Experiment 2, a comb-filtered stimulus was employed as the non-uniform texture. The corresponding type of pdf can be seen in Fig. 1(c), and some examples of the stimuli used are shown in Fig. 4. The pdf consists of a series of equally spaced impulses or Dirac symbols (Bracewell, 1986), multiplied by a constant factor to ensure normalization. In other words, the line orientations are concentrated at a number of discrete angles. Again, we denote the number of orientation modes by N. As N increases, human performance at discriminating such textures from orientationally random ones will obviously decline, because in the limit as N tends to infinity the pdf becomes random. All values of N from 1 up to 12 were used here.

A notch stimulus was used in Experiment 3. The notch pdf is shown in Fig. 1(d) and some example stimuli in Fig. 5. This texture consists of a random pdf except for a sector of angles (or "notch") where the pdf is zero. We assess psychophysical discrimination performance along the dimension of the width, w, of the notch. Clearly, the ability to discriminate the notch array from the random array will improve as w increases. In passing, we note that the term notch, although correct, is perhaps misleading in the light of the percept of this stimulus. In all cases these arrays have a mean orientation which is perpendicular to the notch, and it is this "grain" that is

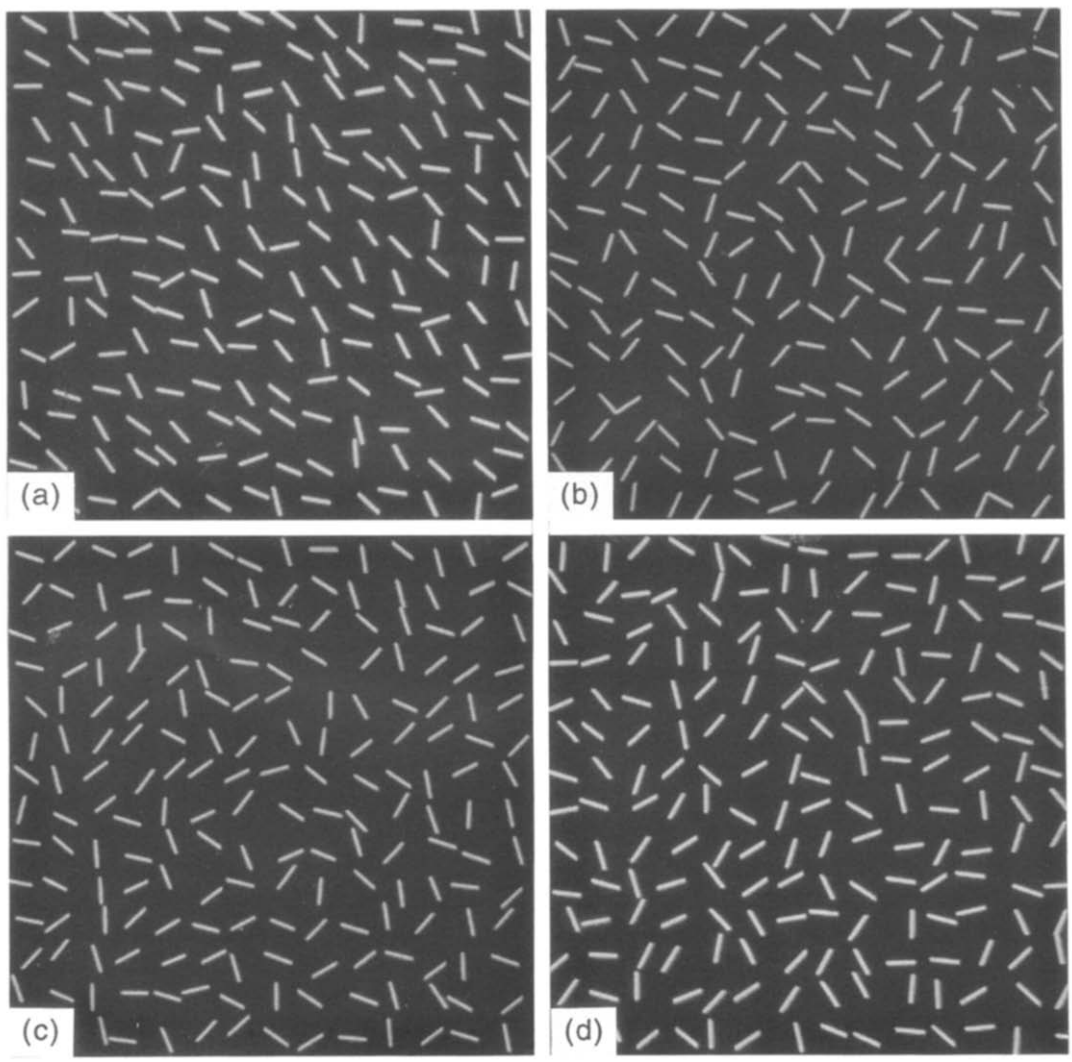

FIGURE 3. Examples of the jittered grid stimuli used in Experiment 1. The arrays were produced using sinusoidal pdfs with frequencies of (a) 1 c/180 deg; (b) 2 c/180 deg and (c) 3 c/180 deg. except for (d), which is orientationally random.

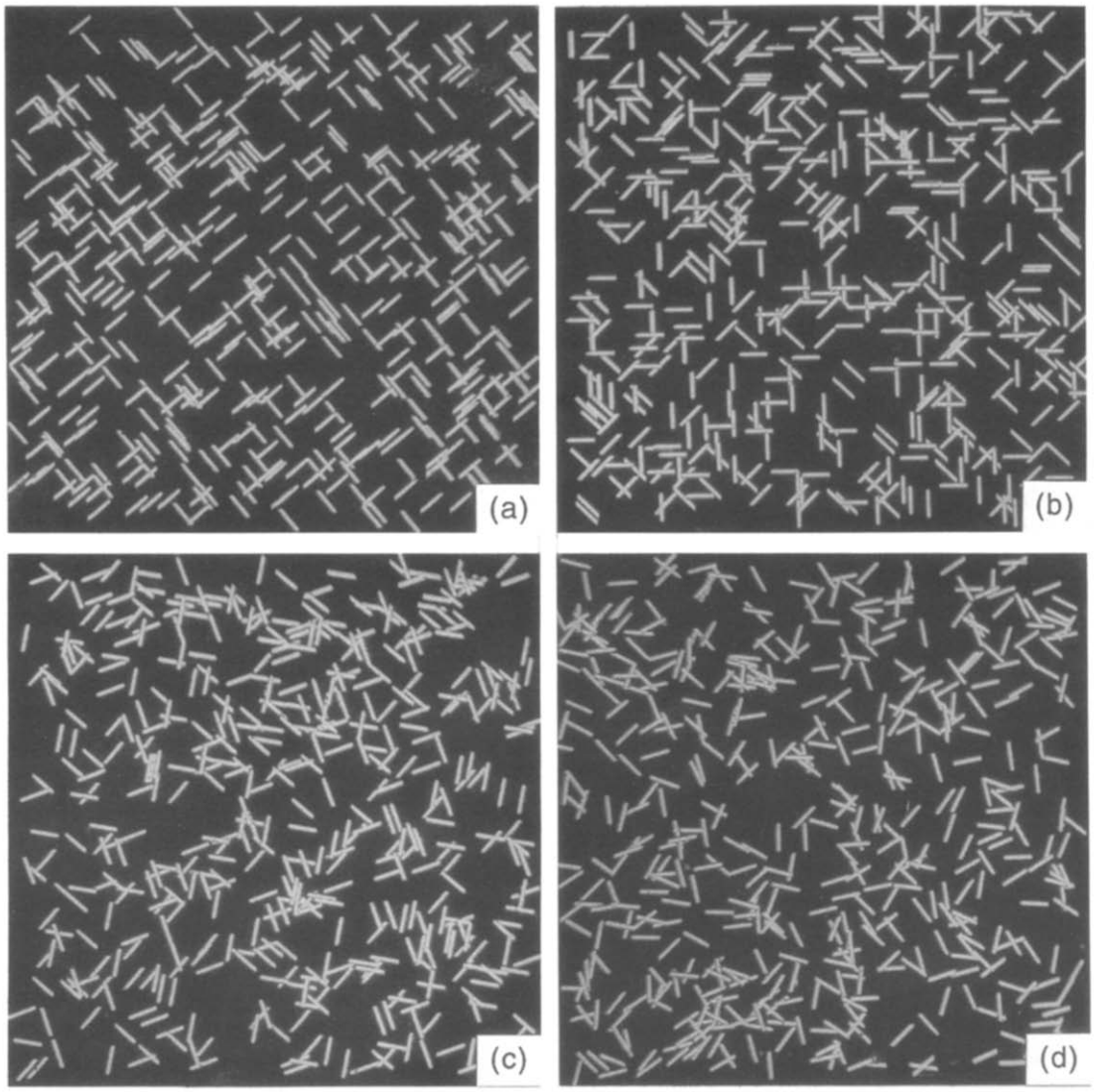

FIGURE 4. Examples of the comb stimuli used in Experiment 2. The arrays on the left were produced using comb pdfs with: (a) 2; (b) 4 and (c) 6 impulse functions per 180 deg, except for (d), which is orientationally random.

perceived by the subject. It should also be realized that the notch width and the amplitude of the pdf cannot be independently manipulated, as the pdf must integrate to a constant. Hence, increasing w implies increasing amplitude. In Experiments 2 and 3 random arrays of 500 lines were again used.

The display screen was viewed from a distance of 57 cm. The line segment micropatterns were all 33 min arc long and 2.2 min arc wide. The arrays were 10.6×10.6 deg square, and were separated by 44 min arc. Because there is an absolute limit to the amplitude of the sinusoidal pdfs, owing to the requirement of non-negativity, it was desirable to maximize observer performance. To do this, we used a relatively long exposure time of 1000 msec to increase the likelihood of performance being above threshold at maximum amplitude, and instructed subjects to use eye movements to do the task, if they so wished. It is conceivable, although perhaps improbable, that these conditions might bring into play some higher-level mechanism employing focused attention, rather than the low-level preattentive mechanism implicated in many classical texture demonstrations. To guard against this possibility, one subject (DK) also performed Experiments 1, 2 and 3 with an exposure time of 105 msec. This is more than adequate to prevent eye movements occurring during the presentation of the stimulus. Fixation was aided by the presence of a white spot displayed before presentation at a position corresponding to the middle of the strip separating the two arrays.

Subjects

In all, six observers were used. Two of the authors, (DK and MJM), served as subjects, and performed all three experiments: they are experienced psychophysical observers. Four undergraduate students, (SMA, RCS, AG and JS) were also used. These four subjects had not previously performed psychophysical experiments. They did not each perform every experiment. All subjects had normal or corrected-to-normal vision.

Procedure

The subjects had to discriminate between the orientationally random and orientationally non-random

arrays in a 2AFC paradigm. They were specifically instructed to attend to the overall orientation distribution of the stimuli. Responses were collected by the observers pressing a key according to which array they thought corresponded to the non-random pdf. Feedback was provided in each experiment. In all four experiments, every subject performed a test block comprising of least 60, and usually more than 100 individual observations in order to familiarize themselves with the task before starting the experiment proper. The method of constant stimuli was employed to produce psychometric functions. In Experiment 1, three equally-spaced levels of amplitude were used, running from 0.1 to 0.5: the maximum possible in our units. In Experiment 2 all modal values from $N = 1$ to 12 were used. In Experiment 3, six notch widths from $w = 0$ to 40 deg were used, apart from one subject, RCS, who, owing to his poor performance in trial blocks, was shown seven widths ranging from $w = 0$ to 48 deg. The side on which the modulated array was presented was chosen such that in half the presentations it was on one side, and half on the other. The full psychometric function was produced by assigning a negative value to the stimulus value where the cue was presented to the left, and a positive value where the cue was on the right. In all cases each point on the full psychometric function represented 30 observations. The psychometric function was thus in terms of proportion where the cue was judged to be on the right. The results were analysed by means of probit analysis (Finney, 1971), which yields a standard deviation (SD), for the best-fitting cumulative-normal fit to the psychometric function. For a bias-free observer, and given that the psychometric function is, in fact, a cumulative normal function, a stimulus level equal to the SD will give a performance level of 84% correct. We take the SD to be our threshold or criterion level of performance. In the case of Experiment 2, the fact that performance declines as the stimulus value (N) *increases*, meant that it was necessary to perform the probit analysis on the psychometric function expressed as a function of the inverse of N. By inverting this SD we obtain an estimate of the value of N at which criterion performance occurred. This is equivalent to considering the stimulus variable to be the angular separation of the impulse symbols.

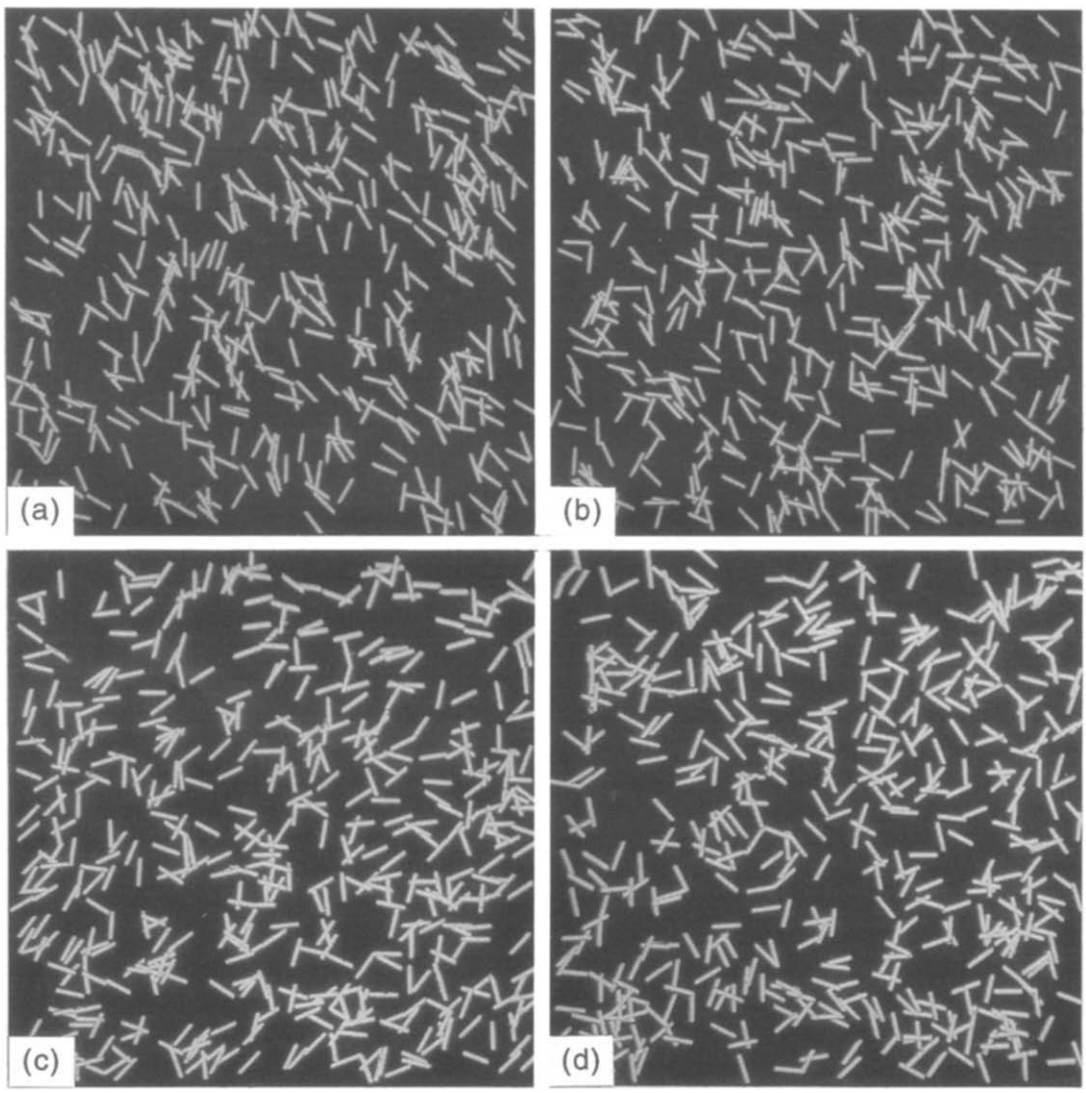

FIGURE 5. Examples of the notch stimuli used in Experiment 3. The arrays were produced using pdfs with notch widths of: (a) 70; (b) 50 and (c) 30 deg, except for (d), which is orientationally random.

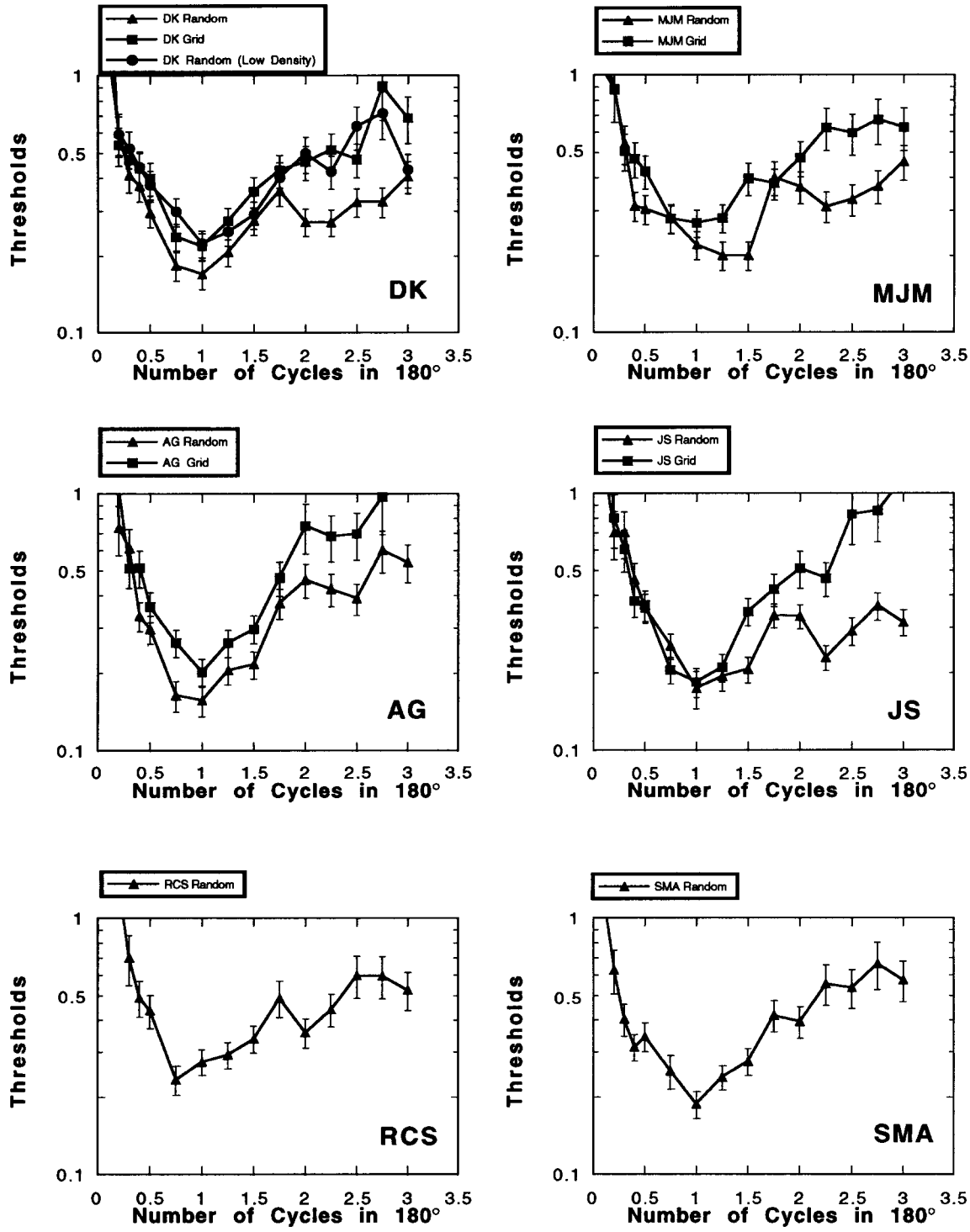

FIGURE 6. Amplitude thresholds for the random and jittered grid tasks used in Experiment 1 as a function of orientation frequency. Subject DK performed an additional random task at a lowered density identical to the grid condition.

Unfortunately, this produced a poor fit to the data, so in this case we simply fitted a logistic function to the data in Fig. 12 and took the 84% correct point as the threshold.

RESULTS AND ANALYSIS

Experiment 1

The results of Experiment 1 may be seen in Figs 6 and 7. The principal feature of the graphs for all six subjects for both grid and random conditions is that there is a prominent minimum at approx. 1 c/π rad, with a steep increase in threshold on either side. So in terms of the orientational frequency the performance of the human visual system in this task can be said to be bandpass. In particular, it is clear that thresholds at 3 c/π are higher than those at 2 c/π which are higher than those at 1 c/π, for a given subject. These three points on the graphs correspond to discrimination performance for the stimuli shown in Figs 2 and 3. The collapse in performance

as the orientation frequency approaches zero is a consequence of the fact that an orientation frequency of zero represents an orientationally random texture—making the task impossible. So in a sense the bandpass result is an inevitable consequence of the restriction that orientation can only change over π rad. It is worth noting that in these graphs thresholds of greater than 0.5 do not denote physically realizable pdfs (owing to the requirement of non-negativity). Thus if the 84% correct level is not reached at a stimulus level of 0.5, the probit routine will produce a non-physical threshold estimate with high standard error. This may well partly explain the noisiness of the data for $N > 2$.

It is of interest that the shape of the OCSF is different from the prediction of the maximum frequency difference discrimination criterion of Voorhees and Poggio (1988). This asserts that texture discrimination is based upon the *maximum* difference between the frequency histograms for any texture attribute. From this one would thus expect the sensitivity to be constant with orientation frequency, as in these stimuli the maximum difference is the amplitude minus the mean level.

We now compare the results for the random arrays and the jittered grid arrays. The thresholds vary with orientation frequency in the same fashion for both types of stimuli, with the minimum being at approx. 1 c/π. In addition, there is a consistent increase in threshold for the grid condition compared to the random condition. This can be explained by the difference in line density between the two conditions. The term "line density" is defined as the number of lines per unit stimulus area. For the displays used in this experiment, the density of the random arrays was 2.5 times the density of the grid arrays. Nothdurft (1985b) has shown that in texture segmentation tasks using grids of oriented lines of the same general kind shown in Fig. 3 the density of lines affects the performance of human subjects: the higher the density the easier the task. In addition, one subject (DK) performed a further random condition where the density of lines was the same as that employed in the grid stimuli. In this case, as may be seen in Fig. 6, the thresholds are almost identical to those from the grid condition. To return to the original motivation for the grid portion of Experiment 1, we can now conclude that the presence or absence of crossovers in these images does not differentially affect their discrimination, and that the mechanism underlying the observed thresholds is orientational in nature.

Figure 7 displays results for the same experiment for subject DK including those conditions where the display time was 105 msec. It is clear that the same kind of small disparity between the random and grid conditions is present at short durations. A comparison between the two spatial arrangements for both 105 and 1000 msec is also shown in this figure. It seems that the opportunity for scrutiny afforded by the long 1000 msec condition produces only a small increase in performance. It is interesting that this only occurs for $N > 2.00$. Possibly this implies that the detection of unimodal textures is *purely* preattentive, whereas the detection of multimodal

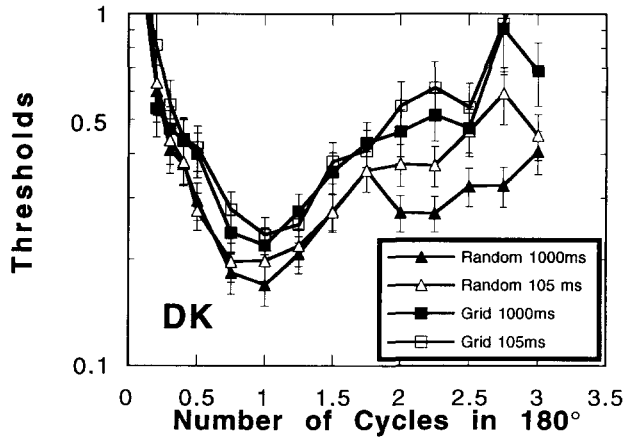

FIGURE 7. Threshold plots for 1000 and 105 msec duration time for both random and jittered grid tasks for subject DK.

textures is slightly enhanced by longer viewing times. In any event, the presence or absence of eye movements does not cause a qualitative change in the results.

The data presented in Fig. 6 are adequate as a description of human performance in this particular texture discrimination task, but we wish to have a more general understanding of how the mechanisms at work here produce the results shown. In order to apply the techniques of Fourier analysis to these results it is convenient to plot the inverse of the threshold (i.e. sensitivity) as a function of the orientation frequency. An across-subjects mean plot for the 1000 msec random condition is shown in Fig. 8. Individual subject OCSFs were also produced. This function is directly analogous to the luminance contrast sensitivity function of the spatial frequency domain (see, for example, Campbell & Robson, 1968), and so we refer to it as the *orientational* contrast sensitivity function (OCSF). Again, we note that the contrast referred to here is the contrast in the sinusoidal pdfs in Fig. 1, and *not* luminance contrast.

The OCSF gives us a description of the sensitivity of the human visual system to sinusoidal modulations in

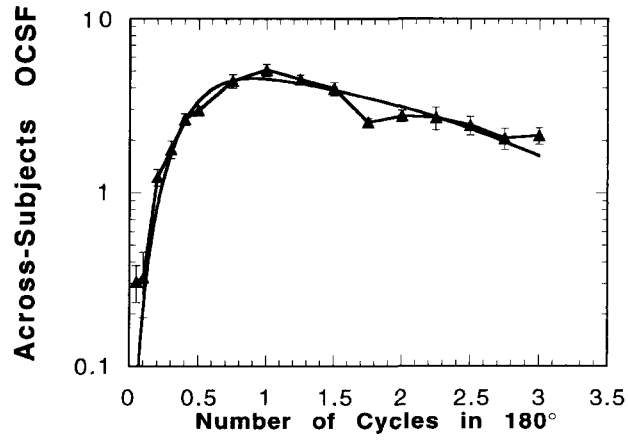

FIGURE 8. The OCSF produced by averaging across individual subject OCSFs for the Random 1000 msec task. The fitted DOG function is also shown.

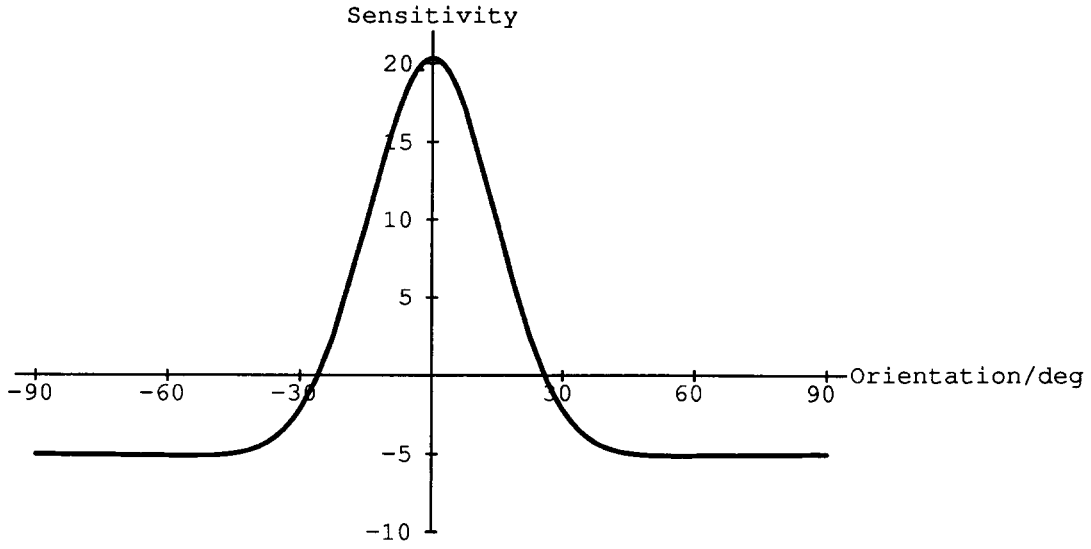

FIGURE 9. An orientational weighting function, produced by means of an inverse Fourier transform performed on the OCSF in Fig. 8. Note that −90 and +90 deg are equivalent points.

the orientational pdf as a function of orientation frequency; we want to express this sensitivity in terms of the orientation itself. To do this, we performed an inverse Fourier transform upon the OCSF, resulting in a plot of sensitivity against orientation. We shall generally use the non-committal term "orientational weighting function" to describe it. A typical example, generated using the across-subjects OCSF is shown in Fig. 9. The method used to perform this inverse Fourier transform is to fit the OCSF with a difference of Gaussians (DOG) function, and then transform this analytically, giving another DOG with different parameters. This method has been used by Robson (1983) and Humanski and Wilson (1992) as a simple method of performing the Fourier transform on experimental data. Although there are no *a priori* theoretical grounds for believing that the OCSF is, in fact, of the form of a DOG, there does not appear to be much systematic structure in the experimental results over and above the bandpass structure, so the fits we employed should be adequate for the purpose of capturing the basic shape of the functions. In fact we also performed a fit employing a difference of exponentials function. The resulting weighting functions in orientation space had similar widths to those we will quote here. The principal difference was the presence of a somewhat implausible cusp at the central maximum. Hence we used the DOG fit.

Because we used a relatively simple function to fit the OCSF, some information about the real structure of the weighting function may have been lost. It was because the individual subjects' data are quite noisy, as can be seen from the thresholds in Fig. 6, that we calculated the across-subjects mean OCSF, which is shown in Fig. 8, together with the corresponding DOG fit. Although the fit is good over most of the orientation frequency domain, there is a small but apparently real dip at about 1.75 c/π. This structure is therefore lost in the fitting process. It is not clear what the cause of this small feature is. One possibility is that the use of non-integral orientation frequencies is introducing additional edge cues in the orientation domain. But it would be expected that this would *improve* subjects' ability to recognize non-random distributions. In addition, such cues should be maximal at half-integral frequencies.

In interpreting the weighting function resulting from the reverse transform it is necessary to consider certain problems caused by the cyclical nature of the orientation domain. The angle of a line is only defined over 180 deg: an orientation value lying outside the selected range will be the same as some value within the range. When one transforms a function into the orientation domain, the result will, in general, not be single-valued. In other words, because the inverse Fourier transformation process does not take the repetitive nature of orientation into account, the resulting function spills outside its limited domain. We have taken the pragmatic approach of adding the over-spill amplitude to the appropriate corresponding points within the central 180 deg region. This produces the single-valued function that one sees in Fig. 9. The effect of the adding procedure is to flatten out the "wings" of the function, meaning that at +90 deg and −90 deg there is no cusp. This operation has the result that the Fourier transform of the orientational function will not be the original OCSF, as one would find if employing normal infinite domains. However, if one convolves the original sinusoidal pdfs with the function, the points composing the fitted OCSF will be recovered. Of course, the convolution takes place cyclically, and over only π rad. Thus one can see that the orientational weighting function simply expresses the information in the OCSF in a different form.

We also attempted to overcome the problem of cyclicity by requiring that the form of the function in the orientation domain be restricted to 180 deg by multiplying a DOG by a box function of width 180 deg. When we transformed this into orientation frequency space

however, the resulting function had an oscillatory structure, due to the box function, which rendered it unsuitable for fitting the OCSF.

The form of the weighting function shown in Fig. 9 is as one would expect from the bandpass nature of the OCSF; that is, the excitatory centre together with an inhibitory surround means that for sinusoidal modulations maximal response will be generated at a particular orientation frequency. In the DOG fitting process the function was constrained such that it must pass through the origin. This means that the resulting filter-like orientational function is balanced. Across subjects for the 1000 msec random condition the mean width of the excitatory portion is 52.2 deg with a standard deviation of 6.2 deg. The across-subjects 1000 msec random OCSF produced a function with a central region 51.7 deg wide. An alternative measure of the width of the function is the width of the function at the mid-point between the maximum and the minimum. The mean full-width for the 1000 msec random condition is 31.4 deg with a standard deviation of 7.3 deg, and for the across-subjects OCSF is 34.1 deg..

We commented above that the weighting function simply reexpresses the information in the OCSF in a different form, and that one can recover the original fitted function by appropriate convolutions. This is the key to the simple model for orientational texture discrimination which we wish to test, (see Fig. 10). If one takes the original sinusoidal stimuli pdfs, at *threshold amplitude*, and convolves them with the orientational filter, then the resulting sinusoids will have a peak to trough distance which is *constant over orientation frequency*. This is a criterion for discrimination in the case of sinusoidal pdfs, and is true by definition, as a consequence of the nature of the Fourier transform. It is an empirical question as to whether or not this criterion will apply for textures produced using other pdfs. This is tested in Experiments 2 and 3. It is worth pointing out that for this model to operate, one must envisage the orientational weighting function being in some sense present at each orientation. That is, the function shown in Fig. 9 must not be thought of as simply being centred around one orientation. Any meridional anisotropy in performance in the original task would cause the weighting function to have different widths or amplitudes at different orientations. Thus the functions we present are *average* weighting functions.

Experiments 2 and 3

To test the model for the notch and comb pdfs we convolved them with the orientational weighting function and examined the peak-to-trough distance of the results. This procedure is shown in Fig. 11 for the notch pdf with a width of 30 deg. The main qualitative effect of this filtering process is to smooth out the sharp edges in the orientation domain; in particular note how the features at +90 deg and −90 deg connect. However, the most important point is the distance from the peak to the trough. We varied the width of the notch and performed the convolution until the modulation was that found for the convolved threshold sinusoids: i.e. the threshold criterion for the sinusoidal case. This notch width is the threshold prediction of the model for the subject whose OCSF was used to produce the weighting function. An analogous procedure was used for the comb stimulus, except that here the variable we changed was the number of impulse functions. As the peak-to-trough distance of the output in these cases is never, except by chance, at criterion, we found our estimate of the impulse function spacing (essentially the inverse of the number of spikes) by interpolation between the two spacings which produce peak-to-trough distances closest to criterion. It should be noted that in Experiments 2 and 3 the dimensions along which performance is measured are comb spacing (or equivalently number of combs), and notch width, respectively.

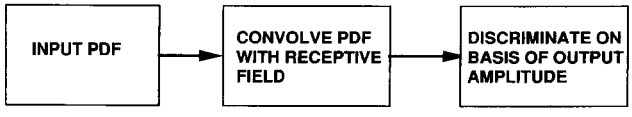

FIGURE 10. The convolution model used to generate threshold predictions in Experiments 2 and 3. Here the weighting function in Fig. 9 is referred to as a receptive field.

The model assumes, in effect, that at some level the visual system has access to information concerning the distribution of texture detail over orientation. Using the pdf as the measure of how this orientational information is arranged as a function of orientation is only one possibility. One disadvantage of this method is of course that the pdf (with which the weighting function is convolved in the model) represents only a distribution averaged over trials—an individual stimulus will have a somewhat different angular distribution. We do not believe that this is a problem in this experiment, because we used a large number of lines (usually 500) and only used orientation frequencies of 3 or fewer c/π. An alternative method would be to use as the orientation distribution for the model the stimulus' Fourier amplitude spectrum as a function of orientation. We believe that this would lead to similar results—but without the tremendous simplifying factor of being able to treat our sinusoidal pdf stimuli as sinusoidal basis functions. We examined two-dimensional Fourier spectra of our stimuli, finding their angular variability to be much as expected from the corresponding pdfs; for example, a notch stimulus had a spectrum with a low magnitude sector of angular width approximately the same as in the corresponding notch pdf.

Subject performance in the comb task as a function of the number of impulse functions is shown in Fig. 12. For all observers performance is essentially perfect up to and including four impulse functions. By seven, however, the response is close to chance level. For many subjects there is a small anomalous second maximum at nine impulse functions. This remains as a noticeable feature after averaging over all subjects for the 1000 msec condition.

Riley (1981) has presented a demonstration in which an "H" figure produced using an $N = 2$ comb pdf [see

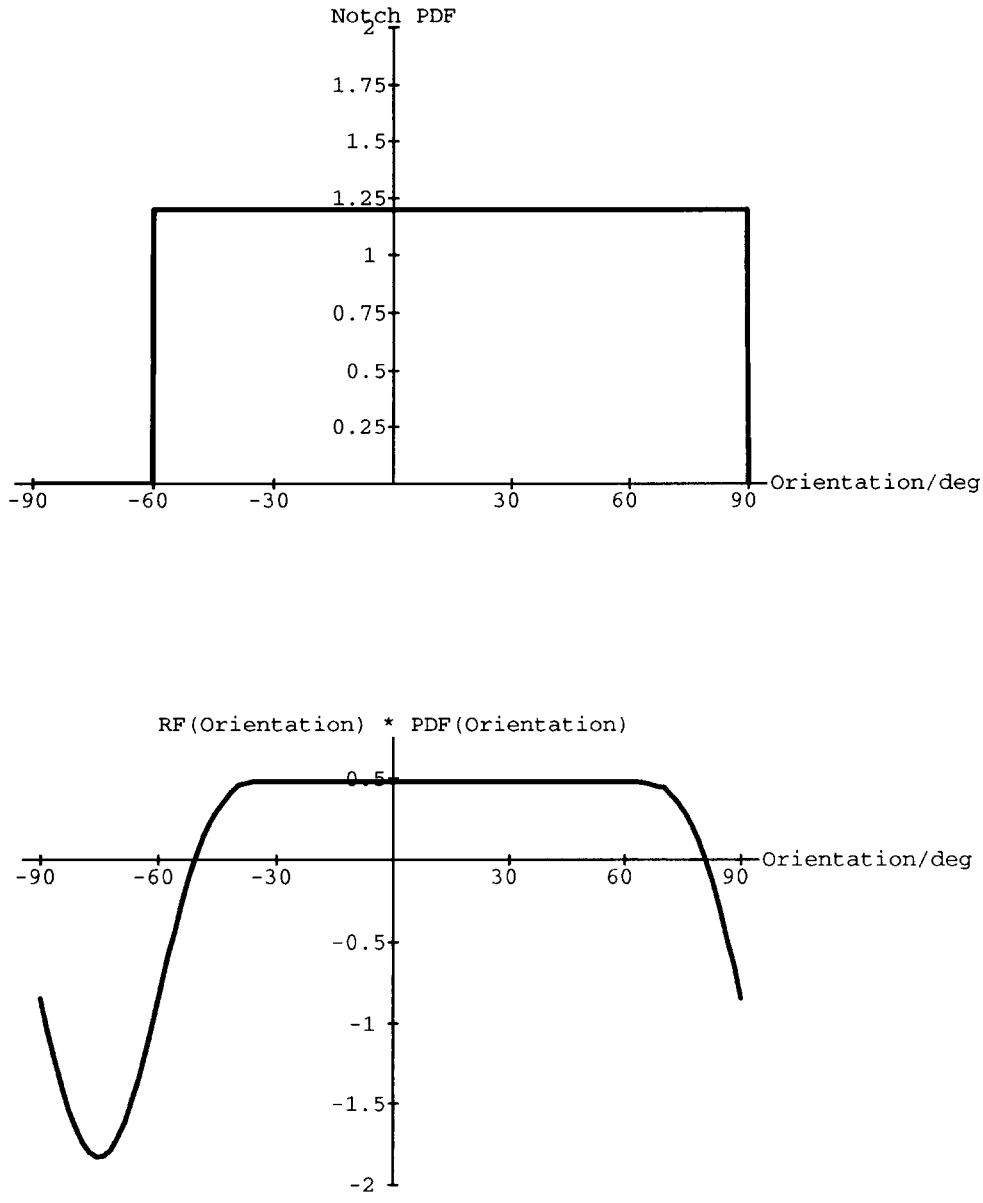

FIGURE 11. The convolution of the weighting function or receptive field with a 30 deg wide notch stimulus centred at −75 deg.

our Fig. 4(a)], appears surrounded by a random texture. The "H" is very hard for observers to segment, showing that in at least some circumstances human performance in discrimination and segregation tasks can differ drastically: performance in our task is near 100% for all observers at $N = 2$. An intermediate case has been presented by Voorhees (1987): this image comprises two simple abutting regions with the same orientational properties used by Riley. Voorhees states that they are not preattentively discriminable, but only on the basis of informal observations: we can segment his image quite easily upon brief inspection.

The experimental thresholds together with the model predictions for Experiments 2 and 3 are shown in Figs 13 and 14. It should be noted that for each subject the convolution prediction is made using the appropriate weighting function generated from that subject's results in the random portion of Experiment 1. Also included are the subject DK's results for a duration time of 105 msec; in this case the weighting function used came from the OCSF produced from the 105 msec task.

First we compare the comb thresholds with the predictions. The mean value are of the same order: 32.2 deg ($N = 5.6$) for the thresholds and 43.5 deg ($N = 4.1$) for the predictions. The prediction supplied by the across-subjects function is higher still: 50.1 deg ($N = 3.6$). The convolution predictions are much more variable (SD for $N = 0.9$) than the thresholds (SD for $N = 0.23$): subject performance in all cases is between 5 and 6 impulse functions. The convolution model does not predict cross-subject variability, but does predict the overall magnitude. One possible explanation for the slightly better than predicted performance on the comb task would be an accelerating non-linearity at some stage in the process whereby the pdf function is mapped onto some representation in the visual system—either before or after the

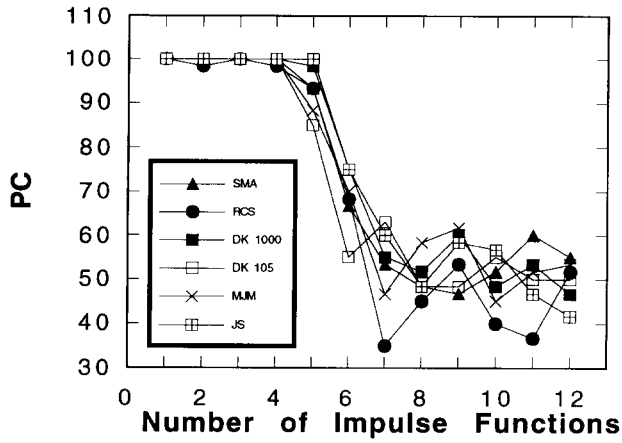

FIGURE 12. Results of Experiment 2. Percentage correct as a function of discrete angles per 180 deg.

process corresponding to the convolution in our model. Other explanations might relate to the presence of constant orientation differences within each array, or the rather frequent occurrence of collinear lines, which could be stimulating the same orientated receptive field.

We discuss the notch results next. The mean prediction from the convolution model is 29.3 deg—very close to the mean of the actual thresholds: 30.7 deg. The across-subjects weighting function generates a prediction of 27.6 deg. The model accounts for the average performance extremely well. In addition, the model produces much of the individual variability seen in the thresholds: the data points lie approximately along the 45 deg line instead of being clustered around it. The source of these differences in subject performance is an open question. It is possible that they reflect divergences in low-level processing, but on the other hand it is conceivable that they indicate attentional or high level differences which presumably would manifest themselves in a similar way in all the tasks employed here. The

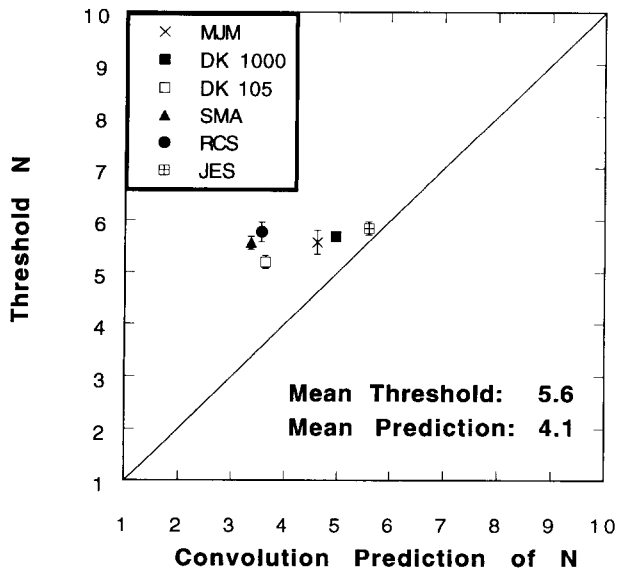

FIGURE 13. The theshold number of impulse functions per 180 deg. from Experiment 2 plotted against the corresponding predictions of the convolution model.

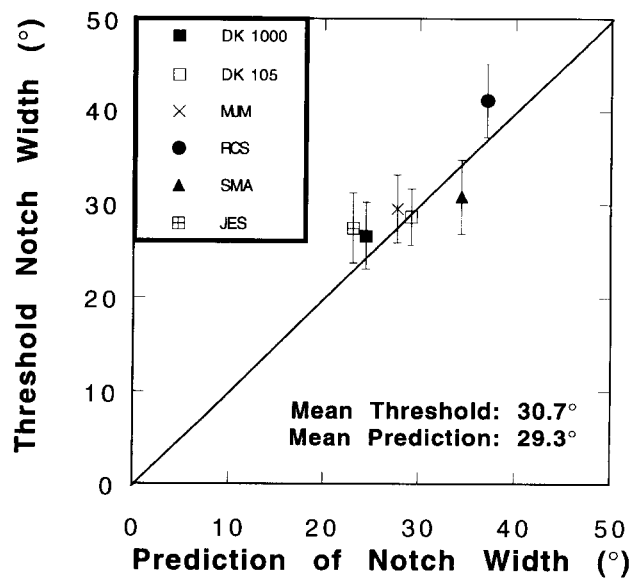

FIGURE 14. The notch width thresholds from Experiment 3 plotted against the corresponding predictions of the convolution model.

convolution model may not account for the subject variability completely—a regression line fitting the points in Fig. 14 is somewhat different from the 45 deg line denoting perfect predictions.

Despite the variability of the quality of the individual predictions in these two experiments, it remains that the mean predictions and the mean thresholds agree well, particularly for the notch stimuli. It should be emphasized that no optimization of parameters has taken place in order to make the model work: the predictions come directly from the orientational weighting function which follows directly from the results of Experiment 1. In both notch and comb results the 105 msec data points fit the model at least as well as the other ones.

DISCUSSION

Summary of results

Before discussing the theoretical consequences of these findings, a brief summary is presented:

(1) Discrimination of sinusoidally modulated textures from random textures is best at 1 c/π rad (unimodal) and deteriorates as frequency increases.
(2) Performance is similar for textures which are positionally random and for those which are grid-based; the small difference is explicable in terms of the line density of the stimuli.
(3) Psychophysical performance for the sinusoidal stimuli can be described by a broad (34 deg half-height full-width) orientational filter.
(4) In similar tasks using notch and comb pdfs, threshold performance occurred, on average, at notch width = 31 deg and 5–6 impulse functions, respectively. These thresholds could be predicted using the independently-generated filter described in (3).

The perception of orientationally unimodal textures

In discussing the significance and status of the results reported here we make the widespread and plausible assumption that texture-based tasks are usually low-level in nature. There are a number of lines of evidence supporting this view. The preattentive and effortless nature of texture segregation seems to imply that high-level processes are not generally used. Prolonging exposure times produces a quantitative but not qualitative improvement in performance (Landy & Bergen, 1991; Bergen, 1991). The ease with which computer modellers have concocted algorithms to perform many texture segregation tasks shows that cognitive processes are not in general required for texture perception. Perhaps the most persuasive evidence for early texture processing comes from the neurophysiological work of Lamme, Van Dijk and Spekreijse (1992). These workers have conducted visual evoked potential studies in Macaque and in Man when exposed to chequerboard texture stimuli. They found a component that could only be due to texture segregation specific activity and localized it to primary visual cortex itself. Bach and Meigen (1992) have also demonstrated VEP activity attributable to texture segregation. There is in addition a fundamental reason why it is probably desirable for a vision system to process textural information at as early a stage as is feasible. Natural textures are inherently intricate in form, and hence require a large quantity of information to be fully specified. To avoid an excessive computational load it would seem efficient to compress this information as much as possible as early as possible. In terms of the orientational content of textures what is probably important is to have some measure of average orientation, orientation variance and overall orientation signal strength, together perhaps with some crude information concerning the shape of the orientation distribution (such as the existence of discontinuities in the orientation content as a function of orientation). The information problem involved in texture processing can be seen in reverse in real-time computer animations (e.g. virtual reality systems). At present these are subject to severe computational bottlenecks and as a consequence textural information is frequently not included in order to speed image generation (Foley, Van Dam, Feiner & Hughes, 1990).

If one accepts the basic premise that the tasks in this paper are being performed on the basis of a modulation in the orientation domain and that this modulation is altered by some process which is described by our weighting function, then the effect of this weighting function is clear: its tuning properties are enhancing the perceptibility of unimodal textures relative to the perceptibility of bimodal and trimodal textures. This is a consequence of the width (average = 53 deg across all subjects and conditions) of the central excitatory portion of the filter. Because the orientational domain in the circumstances of this experiment is limited to 180 deg, the breadth of the filter will mean that almost any two features in the orientational domain will be blurred together by the convolution operation.

There is some other evidence that the visual system in certain circumstances integrates orientational information in a relatively coarse fashion. Foster and Ward (1991) have performed pop-out experiments using oriented lines, and modelled the results using a orthogonal opponent pair of Gaussian filters. The half-height full-width of these functions was approximately 60 deg. In addition, Baddeley and Hancock (1991) have performed a principal components analysis on some natural images, producing bar-detector type functions with widths and shapes similar to those of Foster and Ward. A broad filter will discard much of the detailed information contained in the signal upon which it acts, but in the case of orientational texture information this high orientation frequency information may well not be of great importance.

Why might the visual system be constructed so as to facilitate the discrimination of unimodal textures? It is tempting to answer this in terms of the visual diet that the natural environment provides for the human observer. It seems likely that almost all natural textures are either orientationally random or unimodal when evaluated over sufficiently homogenous image regions. In particular, an orientationally random texture will become unimodal at the retinal level when viewed at an angle owing to the effects of compression—this is important because a majority of the surfaces we encounter in natural scenes are not fronto-parallel. The direction of this induced grain gives important information concerning the tilt of the surface relative to the observer. Unfortunately, although it seems likely that natural processes will tend to produce unimodal distributions, the fact we do not seem to perceive multimodal textures frequently in naturalistic images cannot be taken as definitive evidence of their actual rarity: it is a consequence of the form of the thresholds in Fig. 6 that human observers would not be very sensitive to their presence. We are not aware of any thoroughgoing investigation of the orientational properties of natural textures, although du Buf (1992) concluded that the power spectra of a set of aerial images were almost all unimodal when expressed as a function of orientation. It is of course true that occluding transparent surfaces may produce multimodal orientation distributions. In this case the two overlying surfaces will probably have differences in their properties, in particular their luminance, chromaticity and disparity. These differences may allow texture discrimination to occur separately in the two surfaces. There have been previous attempts to explain the weak discriminability of bimodal textures in terms of the kinds of changes in textures that can be caused by changes in geometrical structure, (Riley, 1981; Marr, 1982). These authors did not, however, propose an algorithm from which the differences in discriminability might emerge naturally.

Interpretation of the weighting function

We now turn to the interpretation of the orientational weighting functions which we have used in this paper. It

is important to recognize that the d.c. level of the function shown in Fig. 9 does not affect the predictions made for performance in Experiments 2 and 3. This is because the pdfs for the comb and notch are all normalized to the same value, and the convolution of a constant, (i.e. d.c.), value with any function of given area will give the same result. Hence, the d.c. level will only shift the convolution result up or down the vertical axis by a fixed amount, but will not affect the degree of modulation. So, it is perfectly possible to consider the filter in Fig. 9 as being entirely positive, with amplitude decreasing monotonically as orientation changes from 0 deg. The half-height full-width of the filter derived from the average OCSF is 34.1 deg. This re-interpretation of a band-pass filter as a low-pass filter is possible because we are dealing with normalized probability density functions, so low orientation-frequency pdfs cannot have high amplitudes. An alternative way of analysing the data presented in this paper, in the light of the actual results, would be to assume that the weighting function is of a Gaussian form and convolve this with all three kinds of pdf used in order to find the best-fitting width.

We are not aware of any evidence for the existence of units in primate primary cortex with the kind of orientational tuning required if the weighting function is to be interpreted as having an inhibitory surround. On the other hand, Macaque neurons (for example), have an orientational bandwidth of median ~45 deg, (De Valois & De Valois, 1990), and a shape of the same general kind as the all-positive interpretation of our weighting function just outlined. The most parsimonious interpretation in the light of current neurophysiological data is thus probably this one, although the inhibitory surround interpretation cannot be ruled out.

Let us then assume that the neural substrate for the detection of orientational modulation in our textures is indeed the ubiquitous orientationally selective neurons in primary visual cortex, and that the orientational properties of these units when processing incoming images correspond in some sense to the convolution model we have used. The question now arises of how the output of these neurons is bound together. One could invoke a 2nd-level filter explanation of the kind used by Moulden (1993) to explain the detection of collinear line segments in an orientationally random background. That is, the initial processing neurons would project to a 2nd-stage array, the members of which would integrate textural information over a region of visual space.

The mammalian striate cortex, however, is not simply a collection of linear filters. There is ample evidence for the existence of lateral connections between orientationally sensitive neurons, (e.g. Ramoa, Shadlen, Skottun & Freeman, 1986; Ts'o, Gilbert & Wiesel, 1986; Ts'o & Gilbert, 1988; Gilbert & Wiesel, 1989). These appear to be both of an excitatory and of an inhibitory nature. There is some evidence both for excitation between neurons with similar orientation preferences and also for inhibition between ones with differing orientation selectivities. The latter has been invoked in some explanations of how orientational selectivity arises (see Ferster & Koch, 1987 for a review). These kinds of interactions can be used to provide an in-place binding mechanism for integrating the texture information, thus obviating the need for a 2nd-stage filter array. The form of these interactions could also play a role in explaining the psychophysical results reported here, as they would tend to enhance the relative amplitude of unimodal textures at the expense of bimodal ones. If this was so, then the orientational filter in Fig. 9 would partly represent an emergent property of these neural interactions. A number of workers have implicated orientation-based grouping processes in psychophysical phenomena (e.g. Zucker, 1985; Field, Hayes & Hess, 1993; Polat & Sagi, 1993, 1994).

Malik and Perona (1990) have used a similar form of cross-orientation and cross-frequency local inhibition in their highly-articulated model of human texture perception. This model predicts the discriminability ranking of texture pairs constructed using different textons. It would probably predict the general kind of discrimination performance we report in Experiment 1.

As mentioned in the Introduction, there are a number of other models of texture perception which employ banks of oriented filters at a variety of spatial scales. These models can generally be applied to arbitrary grey level images. It is far from clear to what extent each of these models could predict our results, although one suspects that any model which blurs in the orientation domain would probably reproduce the *gross* characteristics of our results. The convolution model that we have used is at the algorithmic level: taken directly from psychophysics it describes a mathematical operation that gives human performance in a number of tasks. It is an open question as to how this algorithmic description relates to the essentially hardware-level models which currently proliferate in the literature.

Clearly the convolution model only captures some aspects of human texture perception. It does not address the issues of how texture perception varies with the spatial frequency properties of the component micropatterns, or of how spatial variation of texture properties is perceived. We have investigated aspects of these issues elsewhere (Kingdom, Keeble & Moulden, 1995). In addition the key issue of how textural segmentation (as opposed to discrimination) occurs has not been considered in this paper. It may be, however, that the hypothetical grouping processes that we have mentioned will shed light on this question. Most models of texture segregation generally employ some kind of edge detector over the domain of some function of the output of their first stage linear filters. Excitation between similarly oriented neurons may allow segmentation to occur without this explicit edge extraction stage.

SUMMARY

In this study we have investigated human performance in discrimination tasks involving textures with a variety of orientational properties. A simple orientational filter

model was used to model the psychophysical results. Using this, it proved possible to predict thresholds in experiments employing notch and comb pdfs from experiments using sinusoidal pdfs. This model explained some of the inter-subject threshold variability. The large width of the filter implies that textural orientation coding, at least for the perception of these stimuli, is relatively coarse.

REFERENCES

Bach, M. & Meigen, T. (1992). Electrophysiological correlates of texture segregation in the human visual evoked potential. *Vision Research, 32,* 417–424.

Baddeley, R. J. & Hancock, J. B. (1991). A statistical analysis of natural images matchs psychophysically derived orientation tuning curves. *Proceedings of the Royal Society of London B, 246,* 219–223.

Barlow, H. B. (1982). What causes trichromacy? A theoretical analysis using comb-filtered spectra. *Vision Research, 22,* 635–643.

Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1,* 300–302.

Beck, J., Prazdny, K. & Rosenfeld, A. (1983). A theory of textural segmentation. In *Human and machine vision* (pp. 1–38). London: Academic Press.

Bergen, J. R. (1991). Theories of visual texture perception. In Regan, D. (Ed.), *Vision and visual dysfunction, Vol. 10: Spatial vision.* New York: Macmillan.

Bergen, J. R. & Andelson, E. H. (1988). Visual texture segmentation and early vision. *Nature, 33,* 363–364.

Bracewell, R. N. (1986). *The Fourier transform and its applications* (pp. 69–98). Singapore: McGraw–Hill.

du Buf, J. M. H. (1992). Abstract processes in texture discrimination. *Spatial Vision, 6,* 221–242.

Caelli, T. (1982). On discriminating visual textures and images. *Perception & Psychophysics, 31,* 149–159.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197,* 551–566.

Chubb, C., Econopouly, J. & Landy, M. S. (1994). Histogram contrast analysis and the visual segmentation of IID textures. *Journal of the Optical Society of America A, 11,* 2350–2374.

De Valois, R. L. & De Valois, K. K. (1990). *Spatial vision* (p. 118). Oxford: Oxford University Press.

De Weerd, P., Vandenbussche, E. & Orban, G. A. (1992). Texture segregation in the cat: A parametric study. *Vision Research, 32,* 305–322.

Ferster, D. & Koch, C. (1987). Neuronal connections underlying orientation selectivity in cat visual cortex. *Trends in Neuroscience, 10,* 487–491.

Field, D. J., Hayes, A. & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research, 33,* 173–193.

Finney, D. J. (1971). *Probit analysis* (3rd edn). Cambridge: Cambridge University Press.

Foley, J. D., Van Dam, A., Feiner, S. K. & Hughes, J. F. (1990). *Computer graphics: Principles and practice* (p. 910). Addison–Wesley.

Foster, D. H. & Ward, P. A. (1991). Horizontal-vertical filters in early vision predict anomalous line-orientation identification frequencies. *Proceedings of the Royal Society of London B, 243,* 83–86.

Gilbert, C. D. & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and cortico-cortical connections in cat visual cortex. *Journal of Neuroscience, 9,* 2432–2442.

Graham, N., Beck, B. & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research, 32,* 719–743.

Gupta, S. & Sproull, R. F. (1981). Filtering edges for gray-scale displays. *Computer Graphics, 15,* 1–5.

Humanski, R. A. & Wilson, H. R. (1992). Spatial frequency mechanisms with short-wavelength-sensitive cone inputs. *Vision Research, 32,* 549–560.

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions Information Theory, 8,* 84–92.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232,* 34–43.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature, 290,* 91–97.

Keeble, D. R. T. & Morgan, M. J. (1993). A linear systems approach to texture perception. *Investigative Ophthalmology and Visual Science (Suppl.), 34,* 1237.

Kingdom, F. A. A., Keeble, D. R. T. & Moulden, B. (1995). Sensitivity to orientation modulation in micropattern textures. *Vision Research, 35,* 79–91.

Lamme, A. F., Van Dijk, B. W. & Spekreijse, H. (1992). Texture segregation is processed by primary visual cortex in man and monkey. Evidence from VEP experiments. *Vision Research, 32,* 797–807.

Landy, M. S. & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research, 31,* 679–691.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7,* 923–932.

Marr, D. (1982). *Vision* (pp. 94-95). New York: W. H. Freeman.

Moulden, B. (1993). Functional properties of second-stage orientational filters. *Perception (Suppl.), 22,* 6.

Nothdurft, H. C. (1985a). Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vision Research, 22,* 551–560.

Nothdurft, H. C. (1985b). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research, 25,* 1955–1968.

Polat, U. & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research, 33,* 993–999.

Polat, U. & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research, 34,* 73–78.

Ramoa, A. S., Shadlen, M., Skottun, B. C. & Freeman, R. D. (1986). A comparison of inhibition in orientation and spatial frequency selectivity of cat visual cortex. *Nature, 321,* 237–239.

Riley, M. D. (1981). The representation of image texture. *M.I.T. Artificial Intelligence Lab Technical Report No. 649.*

Robson, J. G. (1983) Frequency domain visual processing. In Braddick, O. J. & Sleigh, A. C. (Eds), *Physical and biological processing of images* (pp. 73–87). Berlin: Springer.

Ts'o & Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *Journal of Neuroscience, 8,* 1712–1727.

Ts'o, D. Y., Gilbert, C. D. & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6,* 1160–1170.

Voorhees, H. (1987). Finding texture boundaries in images. *M.I.T. Artificial Intelligence Lab Technical Report No. 968.*

Voorhees, H. & Poggio, T. (1988). Computing texture boundaries from images. *Nature, 333,* 364–367.

Zucker, S. (1985). Early orientation selection: Tangent fields and the dimensionality of their support. *Computer Vision, Graphics and Image Processing, 32,* 74–103.

Acknowledgements—The work described here was supported by the Image Interpretation Initiative of SERC (U.K.), grant reference GR/F 37124; and NSERC (Canada), grant reference OGP 0121713. Some of the results were presented at the *Association for Research in Vision and Ophthalmology*, 6 May 1993, and have been published in abstract form (Keeble & Morgan, 1993). We wish to thank Roger Watt, Ian Paterson and Steve Dakin for the use of certain computer software.